Analysis Of The Effect Of Hidden Layer Units On Coronary Heart Prediction Using The Radial Basis Functions Algorithm

I Gede Bagus Semara Wijaya[a1], Luh Gede Astuti[a2]

[a]Informatics Department, Udayana University
Badung, Bali, Indonesia
[1]gedebagus097@gmail.com
[2]lg.astuti@cs.unud.ac.id

Abstract

Heart disease is a disease that occurs due to disturbances in the heart, especially when pumping blood so that it can cause death. In Indonesia alone, based on data on disease cases in Indonesia from GLOBOCAN (IARC) in 2012, there were 43.30% new cases of heart attacks and 12.90% of deaths. Therefore, the initial prognosis of heart disease is needed to prevent the risk of coronary heart disease. One thing that can be done is to predict coronary heart disease sufferers by using an artificial neural network radial basis function. This study conducted an analysis of the effect of hidden layer units on the neural network radial basis functions algorithm to predict coronary heart disease sufferers. This study obtained the highest accuracy at 10 hidden layers, namely 85.08%.

Keywords: Heart Coronary Disease, Artificial Neural Network, Radial Basis Function, Dataset, Hidden Layer

1. Introduction

Heart disease is a disease that occurs because of heart problems. Heart attack is one of the most deadly diseases in the world, and one of the diseases that sufferers is Heart Disease. Nearly half of all deaths in the United States and other developed countries are caused by heart disease. In Indonesia alone, based on data on disease cases in Indonesia from GLOBOCAN (IARC) in 2012, there were 43.30% new cases of heart attacks and 12.90% of deaths [2]. The early prognosis of heart disease can help to change the lifestyle in patients who are at a high risk, and also reduce the incidence of heart complications in patients.

Research on heart disease was conducted previously in 2016 by Abdul Rohman, namely research on the comparison of data mining classification methods for prediction of heart disease. The study compared the neural network algorithm, k-nearest neighbor, and C4.5 with the results of testing the three methods using cross validation, confusion matrix and ROC curves, and found that the neural network has the highest accuracy value of 86.06%, then the C4.5 algorithm gets 82.92% accuracy and the k-nearest neighbor algorithm gets 77.58% accuracy. The AUC value for the neural network method got the highest value, namely 0.913, while the lowest was the C4.5 method, namely 0.857 [1]. One method that can be used in data mining is a neural network with a radial basis function algorithm. The RBF network consists of 3 layers, namely the input layer, hidden layer or kernel layer (hidden units) and the output layer. Each hidden unit is an activation function in the form of a radial basis function [4]. Radial Basis Function has the advantage of being faster in conducting training data, and accuracy in estimating [5]. Research on the neural network radial basis function algorithm conducted by Ahmad Setiadi in 2017 regarding the application of the radial basis function algorithm to predict creditworthiness using SPSS neural network 17.0 conducted tests using the configuration matrix method yielding an accuracy value of 0.859 while the ROC curve resulted in an accuracy of 0.627 [3].

Based on several previous studies, this study will use a neural network radial basis function algorithm to predict coronary heart disease. The data used in this study is predictive data for people with coronary heart disease in the next 10 years which are obtained from the Kaggle data base https://www.kaggle.com/dileep070/heart-disease-prediction-using-logistic-regression.

2.   Reseach Methods

In this study, the research method carried out began by analyzing the problems to be raised, then collecting data, then eliminating the missing value in the data, then inputting the tested data into the IBM SPSS Statistics tools, creating scenarios from testing, processing data in SPSS, then copy the test results. The following is a flow chart of the research to be carried out. An overview of this research flow can be seen in the image below:
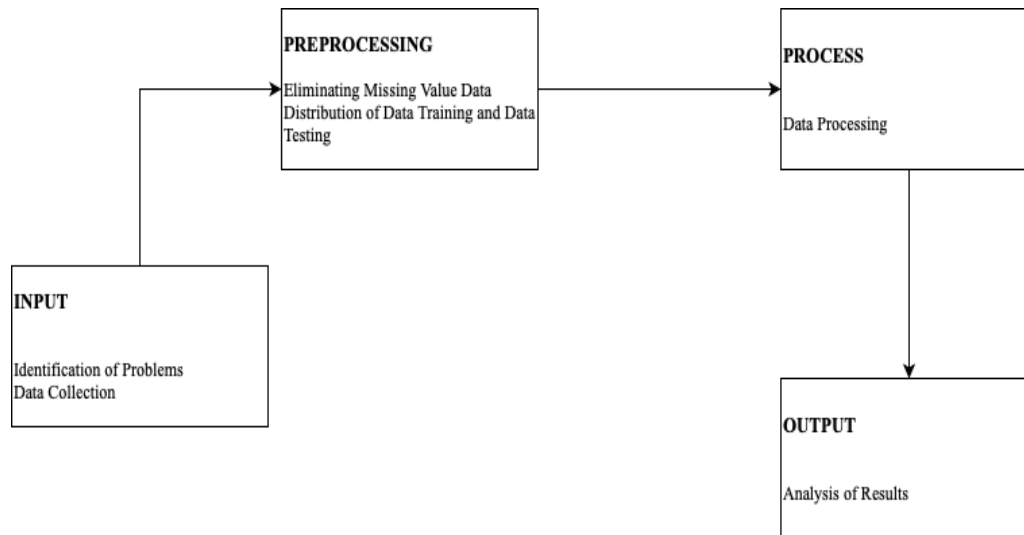
PREPROCESSING

Eliminating Missing Value Data
Distribution of Data Training and Data
Testing

PROCESS

Data Processing

INPUT

Identification of Problems
Data Collection

OUTPUT

Analysis of Results

Figure 1. Research Method Diagram

The following is an explanation of each stage in the image above:

2.1 Identification of Problem

Problem identification is the stage to identify the problem raised and determine the appropriate method to use. The problem raised in this study is to determine the effect of changing the hidden layer on the accuracy of coronary heart disease prediction data using the Artificial Neural Network Radial Basis Function method.

2.2 Collecting Data

This study uses secondary data obtained from https://www.kaggle.com/dileep070/heart-disease-prediction-using-logistic-regression, which is a dataset regarding predictions of coronary heart disease risk in the next 10 years. This dataset has 4238 data records with 16 features and 1 output in the form of a prediction result of the risk of suffering from coronary heart disease and not suffering from coronary heart disease in the next 10 years. The dataset obtained has a missing value and a preprocessing stage will be carried out to overcome the missing value in the data. The following is a description of each feature in the dataset:

Table 1. Domain Dataset Feature

Feature	Domain
Sex	0 - 1
Age	32 - 70
Education	1 – 4
CurrentSmoker	0 - 1
CigsPerDay	0 – 70
BPMeds	0 – 1
PrevalentStroke	0 - 1
PrevalentHyp	0 – 1
Diabetes	0 – 1
TotChol	107 – 696
SysBP	83,5 – 295
DiaBP	48 – 142,5
BMI	15,54 – 56,80
HeartRate	44 – 143
Glucose	40 – 394
TenYearCHD	0 - 1

2.3 Eliminating Missing Value Data

The data in this study have missing value so that data that has missing value will be eliminated or eliminated. To do this, use the SPSS tool by eliminating data that contains missing values so that the data that originally amounted to 4238 data records became 3656 data records after data that contained missing values were removed or eliminated.

2.4 Distribution of Data Training and Data Testing

This process will divide the training data and testing data using a 7: 3 ratio, which is 70% for training data and 30% for testing data.

2.5 Radial Basis Function

Radial Basis Function is one of the Artificial Neural Network methods that are often used. Radial basis functions have good capabilities in non-linear data modeling and models can be formed in one stage, in contrast to Multilayer Perceptron which has to be repeated several times, resulting in more application output faster. The radial basis function algorithm is very useful to solve problems where input data are impure due to noise [6]. Therefore, this study will use the artificial neural network radial basis function method in conducting the data testing process.

2.6 Data Processing

This study uses the Artificial Neural Network Radial Basis Function method, testing is carried out in accordance with the scenario created using data whose proportions are 70% training data and 30% testing data. At this stage, 5 testing scenarios are carried out with 10 iterations.

2.7 Analysis Result

The analysis to be carried out in this study is the effect of unit changes in the hidden layer using 5-25 hidden layer units on the accuracy of the prediction of coronary heart disease using the Radial Basis Function (RBF) method using SPSS tools.

3. Result and Discussion
3.1. Processing Data

At the data processing stage, the Artificial Neural Network Radial Basis Function will be implemented using IBM SPSS tools. In SPSS tools, there is a Neural Network analysis feature to implement the Radial Basis Function (RBF) method. In its implementation, the data used is divided into training data and testing data with a ratio of 7: 3 using 5 scenarios and each scenario using 10 trials. As for the tests carried out in this study, namely:

Experiment	5 Unit Hidden Layer	10 Unit Hidden Layer	15 Unit Hidden Layer	20 Unit Hidden Layer	25 Unit Hidden Layer
	Accuracy	Accuracy	Accuracy	Accuracy	Accuracy
1	86.1%	86%	83.9%	84.3%	83.4%
2	83.9%	84.8%	84.5%	86.1%	84.4%
3	83.8%	84.1%	84.3%	84.8%	83.3%
4	84.1%	85.8%	84%	82.9%	84.4%
5	85.6%	85.1%	85.1%	83.6%	84.7%
6	84.6%	86.7%	85.3%	83.4%	83.7%
7	84.8%	84.8%	83%	84.4%	84.7%
8	85.2%	84.4%	84.2%	85.4%	83%
9	86.5%	83.5%	84.9%	84%	84.7%
10	84.7%	85.6%	84.1%	85.8%	86.5%
Average	**84.93%**	**85.08%**	**84.33%**	**84.47%**	**84.28%**

Table 2. Result of Hidden Layer Change

Based on table 2, it can be seen that the average value of the accuracy obtained for each unit change in the hidden layer by conducting 10 experiments in each scenario, the following are the results of the average accuracy obtained:
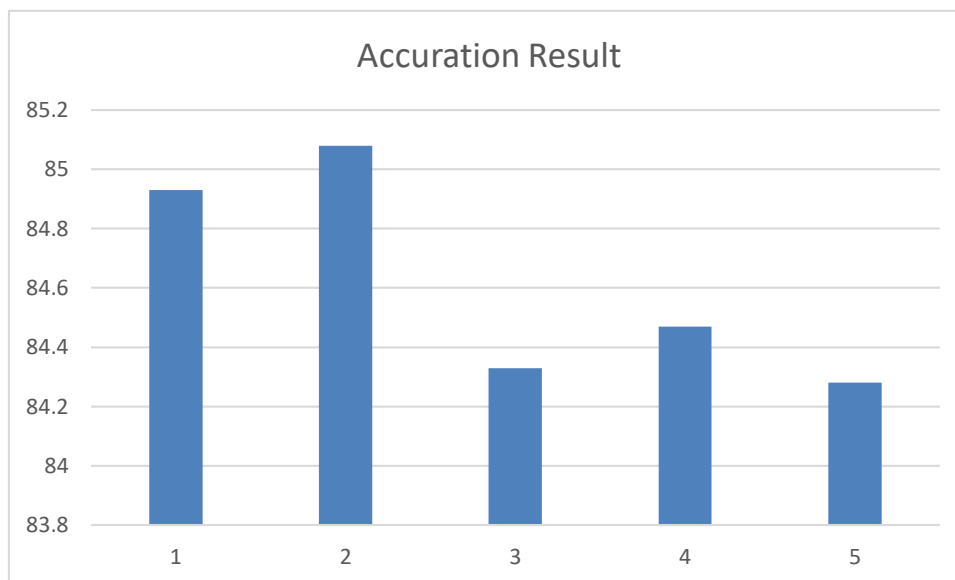
Tabel 3. Accuration Result

Scenario	Accuracy

1	84.93%
2	85.08%
3	84.33%
4	84.47%
5	84.28%

3.2. Analysis Results

The results of the analysis in Table 3 above can be seen in graphical form to better see changes in accuracy in each scenario as follows:

Graph 1. Result

In the implementation results aimed at graph 1, it can be seen the accuracy results obtained from testing unit changes in the hidden layer with 5 scenarios. The highest accuracy obtained is when the hidden layer unit is 10 at 85.08%. Here it can be said that the more and more hidden layer unit changes do not yet have a significant effect on the accuracy results even tend to be lower when compared to the small number of hidden layer units.

4. Conclusion

1. Heart disease is one of the deadliest diseases in the world, and the most common disease is a coronary heart disease. This study uses a neural network radial basis function algorithm to predict coronary heart disease sufferers. The radial base function neural network algorithm has the advantage of being faster in conducting training data, and accuracy in estimating. The data used in this study used secondary data obtained from the Kaggle Dataset. This coronary heart disease prediction data has a total of 4238 data records but because there is missing value in the data which is then eliminated at the preprocessing stage, the data used in the test becomes 3656 data records tested using IBM SPSS tools using Artificial Neural Network Radial Basis Function. On the test.

2. There are 5 test scenarios in this study using changes in hidden layer units in the range of 5-25 units. This study proves that the more and more hidden layer unit changes do not have a significant effect on the results of the accuracy of coronary heart disease prediction, even the accuracy obtained tends to be lower than the number of units in the hidden layer, where the greatest accuracy is obtained when the unit is hidden layer amounted to 10 as much as 85.08%.

References

[1] Abdul Rohman, "Komporasi Metode Klasifikasi Data Mining Untuk Prediksi Penyakit Jantung," *Jurnal Neo Teknika, Vol. 2, No.2, hal. 21-28, 2019*.

[2] Mufti A. Bianto,Kusrini,Sudarmawan,"Perancangan Sistem Klasifikasi Penyakit Jantung Menggunakan Naive Bayes," Citec Journal,Vol. 6, No.1, 2019.

[3] A. Setiadi, "Penerapan Algoritma Radial Basis Functions Untuk Prediksi Kelayakan Pemberian Kredit," *Konferensi Nasional Ilmu Sosial & Teknologi (KNiST)*, pp. 607-612, 2017.

[4] I Dewa G. Budiastawa, I W. Santiyasa, C. Rai A. Pramartha, "Prediksi Dan Akurasi Nilai Tukar Mata Uang Rupiah Terhadap US Dolar Menggunakan Radial Basis Function Neural Network," *Jurnal Elektronik Ilmu Komputer Udayana,* vol. 7, No. 4, 2019.

[5] S. Santosa, A. Widjanarko and C. Supriyanto, "Model Prediksi Penyakit Ginjal Kronik Menggunakan Radial Basis Function," *Jurnal Pseudocode,* vol. 3, no. 2, pp. 163-170, 2016.

[6] Amrin, "Perbandingan Metode Neural Network Model Radial Basis Function Dan Multilayer Perceptron Untuk Analisa Risiko Kredit Mobil," vol. XX, No. 1, 2018.